The Many Use Cases of Standardized Data and Metadata

Sanjiv Ramalingam, Biogen, Cambridge, USA

ABSTRACT

Study Data Tabulation Model (SDTM) datasets are predominantly Clinical Study Report (CSR) driven and primarily created to support analysis dataset creation. The data flow is typically unidirectional. The SDTM group has pioneered the creation of SDTM datasets enabling creation of SDTM datasets within days of First-Patient-In (FPI) and automatic refreshes enabling wider use of standardized data across the organization. With this new way of working, the SDTM group is able to serve and enable efficiencies across multiple functions such as Data Management, Data Standards and Governance, Biomarker and Clinical Operations than just Statistical Programming. The use cases for each of these groups have been discussed in this paper.

INTRODUCTION

There has always been a demand as well as an underlying basis for access to standardized data as well as metadata within Pharmaceutical/Biotechnology companies. The direction of data flow once data is collected has typically remained unidirectional as illustrated below. The red arrow indicates flow of data from the SDTM group.

In 2018, the company management made a strategic decision to switch from an outsourced model of SDTM to an insourced model. Ever since, the SDTM group has made great strides in implementing a metadata driven SDTM automation process[1] that has enabled the group to create SDTM datasets within days of FPI. Strong internal collaborations enabled us to create data pipelines to download data (EDC and External) every day and provide everyday SDTM refreshes as well.

Additionally, a Metadata Library reflecting a collection of metadata consisting of select fields from several legacy and ongoing study database specifications was integrated with study operational information and equipped with a robust search functionality was implemented at the company wide level. The library called BRaMA[2] (Biogen's Raw Metadata Access) is integrated with SDTM annotations and it's use cases have also been discussed. At Biogen, we have now moved from the unidirectional data flow to the data flow below for the use of standardized data. CSDRIVE is an initiative of Data Management to manage all source data flow and architecture and provide Clinical Data Analytics capabilities.
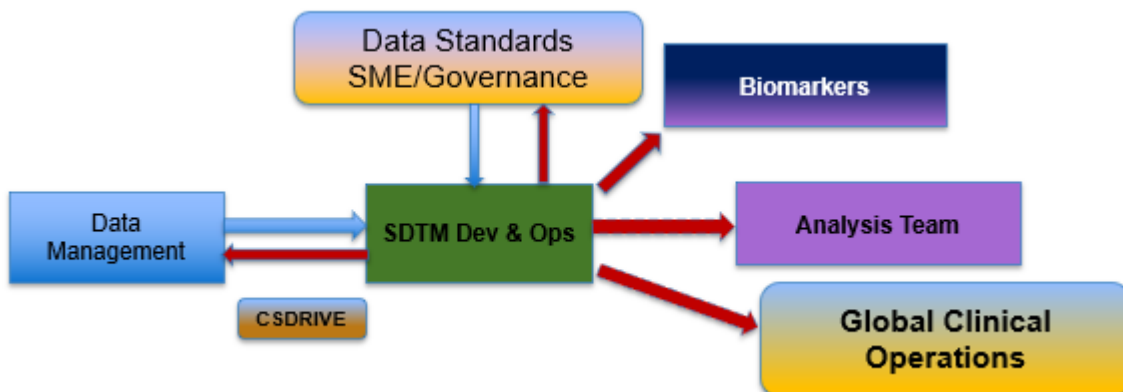
The scope of the paper will focus on the multiple functions the SDTM group has been able to serve.

METADATA APPLICATIONS

A Raw Metadata Library can be a useful resource for multiple stakeholders in Biometrics at various stages of the clinical trial process. At Biogen, BRaMA assists Data Management with the design of the database to reference prior assessments and is now integrated with the EDC specification process. Prior to the creation of such a library the database specification creators had no means to look for assessments from a pooled library.

BRaMA has immensely benefited the SDTM programmers by automating the population of SDTM mappings for new study case report form metadata by integrating with annotations already present in BRaMA of Global mappings and leveraging past study annotations. Governance requests to SDTM SME have also fallen as a result of having such a metadata library and has become a resource for the SDTM SME.

BRaMA assisted the Biomarker team to address implementation gaps for the new In Vitro Diagnostic Medical Devices Regulation (EU) (IVDR). Study teams were asked to identify measurements that are being conducted with EU patients and if within the scope of regulation to determine if a CE marking is required. CE stands for Conformité Européenne (French), which means European conformity. A CE marking implies that the legal manufacturer has assessed the device, and that the device meets the General Safety and Performance Requirements under the IVDR 2017/746. Having a library of assessments by study made the task programmatic and easier to manage than create the list of assessments manually.

GLOBAL CLINICAL OPERATIONS

The company is committed to Diversity and Representation in its Clinical Trials. A dashboard[3] has been developed by Global Clinical Operations to view the US participant demographic distribution in Biogen trials (race, ethnicity, age, gender) and compare to the epidemiology of disease areas. The source data for the dashboard relies on SDTM data from individual studies and ADS is tasked with providing the source data. The GCO Underrepresented Population Dashboard is reliant on teams updated SDTM data and study teams use the dashboard as the source of truth.
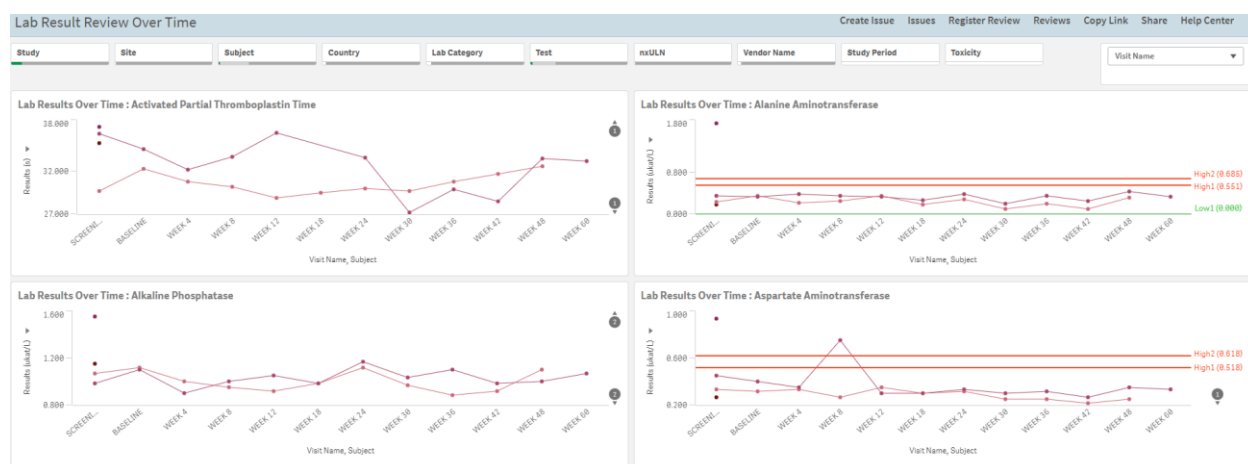
A cross-functional automated data flow to support Under-Represented Population Initiative is illustrated below.
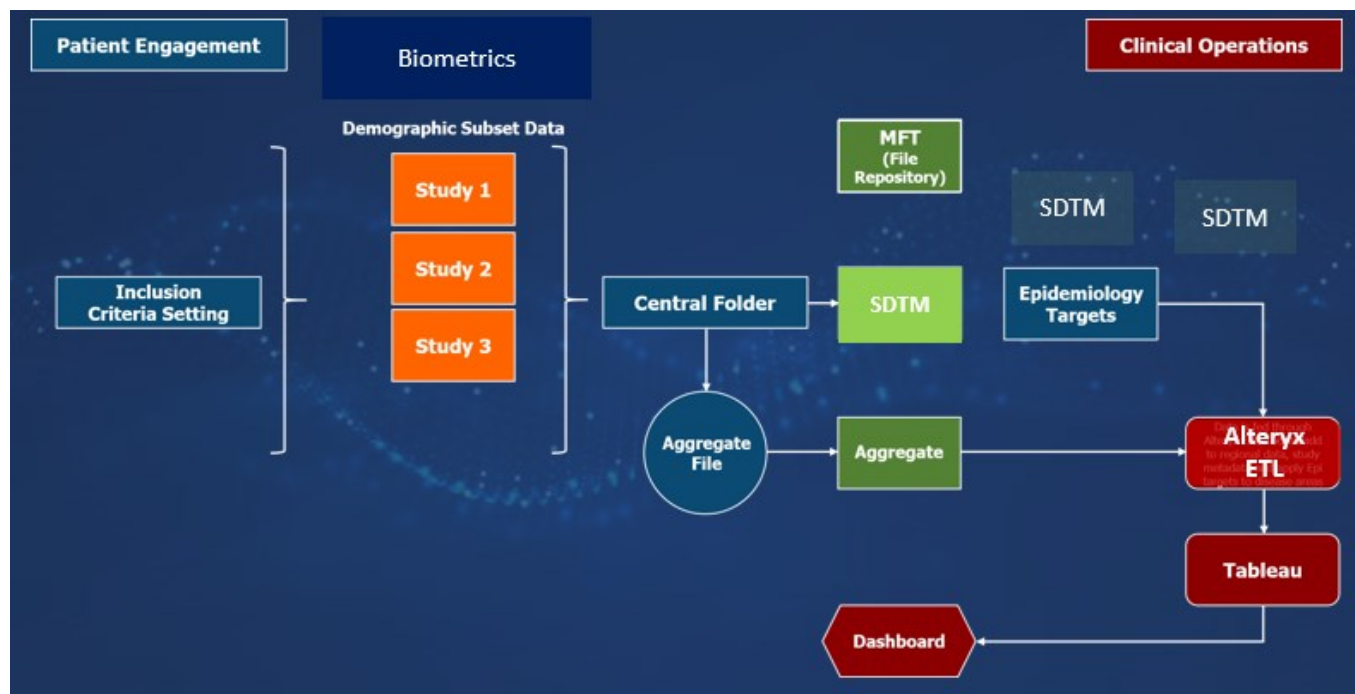
The diversity targets are determined by Epidemiology and tracked by Clinical Operations at the study level. The SDTM group ensured that an SDTM like dataset inclusive of demographic information was created for every study and automatically run once a week based on weekly raw data extracts and posted to a central folder. A separate program was created to anonymize and pool all the individual study demographic data. The aggregate file is then used as a source to create an interactive dashboard using Tableau.

DATA MANAGEMENT

The SDTM group is proud to support an initiative within Data Management to reduce database lock(DBL) timelines between LPLV (Last patient last visit) and DBL by supporting data cleaning initiatives using data visualization and analytics(DV&A). The DV&A relies on standardized data as a source that is refreshed on a day-to-day basis. The cross-functional groups collaborated to create a data pipeline that facilitated seamless movement of data between the groups. Within Clinical Data Management, the Clinical Data Review (CDR) Team reviews aggregated outputs and visualizations to identify data anomalies real-time/upstream utilizing SDTM domains and data model with SDTM data modelling serving as the single source of standardized data. SDTM domains are used for developing outputs for cleaning data based on IDRP(Integrated Data Review Plan) requirements . Utilizing standardized data model brings programming efficiencies (outputs programmed once and validated do not need to be redone) as well data cleaning efficiencies. Clinical data and operational metrics are synchronized/aligned for data managers to effectively manage the studies and near milestones the individual study refresh frequency will be increased to more efficient cleaning; ultimately aiding the Database Lock cycle time reduction and overall E2E clinical study timeline to reporting/submission. Snippets of outputs that use SDTM datasets in real time(everyday basis) that aid in anomaly detection are shown below :

Lab Result Review Over Time

Lab Results Over Time : Activated Partial Thromboplastin Time

Lab Results Over Time : Alanine Aminotransferase
High2 (0.685)
High1 (0.551)
Low1 (0.000)

Lab Results Over Time : Alkaline Phosphatase

Lab Results Over Time : Aspartate Aminotransferase
High2 (0.618)
High1 (0.518)

Lab Numerical Results by Visit Name

Test	UNSCHEDULED - DAY -64	UNSCHEDULED - DAY -52	UNSCHEDULED - DAY -42	UNSCHEDULED - DAY -40	UNSCHEDULED - DAY -35	SCREENING	UNSCHEDULED - DAY -34	UNSCHEDULED - DAY -30	UNSCHEDULED - DAY -29	UNSCHEDULED - DAY -28	UNSCHEDULED - DAY -27
Activated Partial Thromboplastin Time (s)	-	-	-	37.200	32.800	33.159	-	28.000	30.400	-	-
Alanine Aminotransferase (ukat/L)	-	-	0.468	0.184	-	0.340	-	-	-	-	-
Albumin (g/L)	-	-	44.000	51.000	-	36.122	-	-	-	-	-
Albumin/Creatinine (mg/mmol)	-	-	-	-	-	4.749	-	-	-	-	-
Alkaline Phosphatase (ukat/L)	-	-	0.935	1.470	-	1.220	-	-	-	-	-
Aspartate Aminotransferase (ukat/L)	-	-	0.468	0.217	-	0.357	-	-	-	-	-
Basophils (10^9/L)	0.010	0.060	-	0.070	-	0.049	-	-	-	0.040	0.060
Basophils/Leukocytes (%)	0.200	1.200	-	1.200	-	0.804	-	-	-	0.500	0.900
Bicarbonate (mmol/L)	-	-	28.000	24.000	-	24.991	-	-	-	-	-
Bilirubin (umol/L)	-	-	5.130	10.260	-	9.886	-	-	-	-	-
Calcium (mmol/L)	-	-	2.300	2.450	-	2.407	-	-	-	-	-
Chloride (mmol/L)	-	-	101.000	99.000	-	102.144	-	-	-	-	-
Creatinine (umol/L)	-	-	102.544	100.776	-	4,508.346	19,000.000	-	-	-	-
Direct Bilirubin (umol/L)	-	-	-	-	-	4.554	-	-	-	-	-
Eosinophils (10^9/L)	0.040	0.120	-	0.160	-	0.134	-	-	-	0.070	0.240

CONCLUSION

The company has undergone a paradigm shift in its utilization of standardized data, moving beyond conventional data flow. Each successful implementation becomes a compelling use case and demonstrably expands the spectrum of use of standardized data. We envision further exploration of this approach to harness the full potential in biopharmaceutical research and development.

REFERENCES

[1] Streamlining SDTM Development: How Standards and Metadata-Driven Programming can Eliminate Double Programming. Sanjiv Ramalingam, SDTM Development and Operations, Biogen Inc. CDISC US Interchange 2023

[2] DIY: Building a Raw Metadata Library with Integrated SDTM Annotations. Sanjiv Ramalingam and Shriram Sivaraman, SDTM Development and Operations, Biogen Inc. PhUSE 2023

[3] Demographic Distribution Comparison of Biogen's US Multiple Sclerosis Clinical Trials to the Underlying Disease Population: A Decade-Long Retrospective Analysis. I. Mehta, A. Palleschi, L. Kniola, A. Jain, S. Ramalingam, K. Wilson; Biogen, Cambridge, MA.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Sanjiv Ramalingam
Company: Biogen
Address: Binny Street, Cambridge, MA
Email: sanjiv.ramalingam@biogen.com

Brand and product names are trademarks of their respective companies.